

GROS

GROS ONC STRATEGY

ONS(ONC(SC))00/02

1. The purpose of this paper is:
 - to describe the GROS ONC and CCS Strategy;
 - to explain the differences between the ONC and CCS in Scotland and in England and Wales;
 - to allow an assessment of the ramifications; and
 - to seek comment.

The GROS ONC and CCS Strategy

2. The GROS strategy is the same as ONS; that is to run a survey of about the same size as England and Wales on a *pro rata* basis, using a similar design. The CCS will collect data that can be used in the ONS ONC estimation processes including matching, DSE estimation, Design Group age/sex estimation, national level population estimates, LAD level estimation and synthetic imputation.
3. While there are differences, the GROS strategy is limit these to those we deem necessary. Differences in detail between the methods employed in Scotland and England and Wales therefore can be seen to arise from differences in local government administration arrangements, geography, Census geography, and the quality of administrative data.
4. These differences are considered below under the following headings:
 - 1) Design Groups
 - 2) Survey Design Efficiency
 - 3) CCS Field Methodology
 - 4) Sample Design
 - 5) Contingency
 - 6) Imputation of Synthetic People

Design Groups – also known as Estimation Areas (EAs)– based on Health Board Areas

5. Estimation Areas (EAs) in England and Wales are based on groups of Local Authority Districts (LADS). England has 101 EAs covering 376 LADS. Scotland has 8 EAs based on Health Board Areas. Except for areas in the old Strathclyde Region around Glasgow, Council Areas nest neatly into Health Board areas; 28 Council Areas nest perfectly, 4 Council Areas are split by Health Board Area boundaries.
6. The Estimation Area design for Scotland is at Annex A. Estimation Areas in England and Wales have populations from 402,000 to 1,013,000 people; in Scotland they vary from 427,000 to 911,000.
7. Population estimates are more certain at HBA level because migration between HBAs is measured directly, but only apportioned to Council or split Council Area level. If the EAs

reflect HBAs, GROS has more certain HBA population estimates to compare with the ONC EA population estimates at the quality assurance stage of the ONC.

8. Like ONS, GROS will have a two-step estimation process. The first step is to estimate age/sex totals for Estimation Areas. The second step is to estimate Local Authority Area (or part council area where they were split for the Scottish design) age/sex totals – the proposed method is described in ONS (ONC(SC))00/03.
9. The proposed Estimation Area design allows GROS to retain the simple option of summing LAD (including split council area) level ONC estimated age/sex totals to Health Board Area age/sex totals. It is at HBA level that some administrative data sets are most accurate. Work is continuing at GROS to assess the quality of the available administrative data sources. Also it must be remembered, that the size of the CCS in Scotland will not allow as accurate estimates of national totals as those for England and Wales.
- 10. The GROS Estimation Area design reflects Scottish administrative geography and data sets and provides the best design for QA purposes. We believe that it requires no extra processing development work from ONS.**

Survey Design Efficiency

11. We propose to make estimates for 36 council and part council areas (I am calling this group of 36 divisions ‘**Baileries**’) - rather than the 32 council areas we would have used, if the HBA consideration was omitted. We have an average of 4.5 Baileries per EA; England has an average of 3.7 LADS per EA.
12. In response to the increased LAD equivalent divisions - Baileries, we have planned to increase the sample size, as each interviewer will cover more households. Our first stage sample nonetheless is proportionately approximately the same as England and Wales (ONS has 4,000 1st stage EDs for a population of 52.4 M, compared to Scotland of 400 for a population of 5.1 M).
13. GROS will also ensure that each Bailery has at least 3 sample points (At least 1 for every HTCI group). Thus, Orkney, Shetland, Eilean Siar and the 8 split council areas will all have a minimum of 3 1st stage ED samples. Currently these samples are at the expense of sample points elsewhere - chosen from the most over represented HTCI group in other Baileries within the same EA. Table 1 shows overall sample details for Scotland and England and Wales.
14. Splitting 4 council areas into 8 Baileries inevitably decreases the average population of our LAD level areas. In particular, one area, Chryston, has a population of 16,000. This can be compared to the the smallest whole council areas in Scotland of Orkney and Shetland with 20,000 and 23,000 people respectively and, in England, the Isles of Scilly, City of London and Teesdale with 2,000, 5,000 and 25,000 people respectively.

Table 1: Population and CCS sample size in Scotland and in England and Wales

Country	Population	1 st stage sample size	No of interviewers	hhlds per interview	Sample of post codes	Sample of households
Scotland	5,120,000	400 EDs	400	100	2,500 **	40,000
England and Wales	52,427,900 *	4,000 EDs	3,600 **	83 **	20,000	300,000 **

* mid 1998 population in 'Population Trends', Winter 1999.

** estimate – can't be known till sample selected.

15. GROS is committed to the HBA design group approach and accepts responsibility for ensuring that any knock on effects in the system of using split Council areas are properly addressed. GROS currently plans to ensure a minimum of 1 1st stage sample point for each HtC strata for each council and part council area. This can be achieved by either:

- Increasing the sample size by adding a new sample point where necessary to meet this objective; **or**
- Moving a sample point within the Estimation Area from the most over-represented council by HtC group.

16. Any costs to meet the first option would be small, while the second option could lead the sample to favour Orkney, Shetland and the Western Isles rather than Inverness perhaps and peripheral areas around Glasgow over Glasgow City. We would welcome advice on adjusting the sample location versus adjusting the sample size.

Field methodology

17. In detail, there are a number of differences in field methodology between GROS and ONS. However, the overall approach of the CCS projects at ONS and GROS is the same. We both require team working and have proposed strict quality checking in the field with a management and support structure which is robust enough to deal with personnel problems, non-contact and difficult area working. The two projects are liaising closely on these matters.
18. Each GROS interviewer will cover whole postcodes in randomly selected postcodes within a 1st stage ED. In the current design, each will cover more than 95 households, but on average about 100 households.
19. We will support the interviewers' work using 2 standbys per Team Manager, a Team Manager's Assistant, and finally a fellow interviewer if required. The dispersion of the sample in Scotland makes it difficult to create practical pairing arrangements before the event – the Team Manager will create pairing if and when necessary. (It should be noted that the Team Manager Assistants' primary role is to carry out various quality checks.)
20. We will deal with variability between workloads by having in place a piece rate and an hours-based bonus scheme in cases where contact is difficult. The law now requires that employees doing unmeasured work must keep hours worked logs - so no extra administration is needed for either of these schemes.

21. In rural areas, our rehearsal work indicated that customised arrangements may be required - to some extent. Therefore I can not offer a final methodology for our rural areas, but options available are to:
 - Reduce the workload (i.e. no of households);
 - Provide map-reading escorts (difficult to read and drive);
 - Increase the pre set hours;
 - Have 2 interviewers in one area.
22. We believe that the provisional field results from Scotland and England and Wales indicate that, to ensure the CCS meets its targets, further improvements, especially in quality control, training focus and efficient working hours spread are required. GROS will report on its evaluation and recommendations along with the CCS team at ONS.
23. The CCS team has recognised the problems of scaling up the survey. There are also problems of dispersion since the CCS rehearsal covered 17.5% of the rehearsal area and the 2001 CCS will cover only about 1.5% of the country. In Scotland we have met some of these problems with the single workload proposal, but we will also have to further consider our field management structure – when the sample is finalised – and training of isolated interviewers.
24. There are many quality issues in the MIS and survey control yet to be resolved. ONS and GROS teams are working closely on these issues and are coming to very similar conclusions. There is little doubt that we will have very similar quality controls in place in the field in 2001.
25. **Sample dispersion in Scotland makes it difficult to provide cost effective, practical, partnership workload areas comprising more than 1 cluster. Other considerations and the results of the rehearsal provide evidence that in Scotland, a single cluster workload design is cost effective, efficient, and practical.**

Sample Design

26. ONS have run simulations to measure the efficiency of their sample design. GROS started similar work but ran into problems. We are hopeful that we have navigated around most if not all such problems and some correctly formatted data sets have been given to ONS to check.
27. However, I start with the premise that, if we were to follow the ONS sample design, we would achieve similar estimate errors for age/sex groups for each Estimation Area. We may have a more homogenous population than England and Wales, but perhaps at the Estimation Area level this is not significant. We have some Bailerries with relatively small populations, but the average population of a Scottish Bailery is about the same as the average population of a LAD in England and Wales.
28. The GROS proposal is that, at each 1st stage sample point enough postcodes are sampled to create a single workload. It provides a simplification of the fieldwork, to be offset against loss of statistical efficiency compared to the ONS proposal. It:
 - increases the sample size (in terms of postcodes and households covered) but reduces relative statistical efficiency – theoretical accuracy will be the same or perhaps marginally increased;
 - reduces the travel time and distance for interviewers and therefore costs - in a dispersed sample;

- thereby concentrating the interviewer's efforts in one local area; and
 - ensures that, at each 1st stage sample point, an interviewer's resource is not wasted by giving them an undersized workload.
29. Therefore, GROS is exploring a design of randomly sampled postcodes within an ED until a threshold of 95 households has been crossed. Starting with the premise in paragraph 27, I believe this design meets the 2nd stage random postcode selection criteria preferred by ONS and provides cost efficiencies against a slightly larger, but probably marginally more inefficient, sample.
- 30. GROS believes we can simulate such a design (that is a cluster of randomly selected postcodes within a 1st stage to form a single workload) and that there are no significant ramifications for down stream processing development.**

Contingency

31. Contingency is covered by other documents being provided to the ONC SC for this meeting. Therefore, I am simply covering the strategy here.
32. If the CCS produces a result in an EA that the QA process identifies as unlikely, both GROS and ONS would look to 'borrowing strength' as a method for reconciling the QA failure. I understand that, like GROS, ONS will have some regard for administrative data at the stage of quality assuring EA age/sex populations, although the need for 'borrowing strength' may also be identified by determining demographic ratios such as dependency or sex ratios.
33. GROS is committed to a similar approach to that adopted by ONS, although the exact QA approach, covering DSE/Census ratios, EA estimation population totals and age/sex EA estimates, requires further detailed specification. It is also the case that each Census Office may apply a different weight or value to each of their comparative administrative data sets, particularly where such data appears to provide a viable independent estimate.
34. GROS feels that there may be circumstances where the results will require something other than the precise process of borrowing strength described in the ONS papers. We suggest that further work is required to agree what flexibilities there are within the QA procedure and subsequent action. This may be best done by a joint review of practical examples.
- 35. GROS is aware of the need for careful presentation of the results of any adjustments made to the Census data that may fall outside the specific pre-determined ONC methodology. We suggest that further consideration be given to dealing with such an eventuality in any of the 3 Census Offices.**

Imputation of synthetic people

36. If administrative data is used to control population totals for age sex groups at the EA level, the question arises whether the characteristics of the people missed by the Census can be used to impute synthetic people etc. If you have rejected the data for estimation, can you use it for imputation?

37. There appear to be two options for GROS, either
- a: Accept that the CCS does provide useful information about the characteristics of the missing people and use this even when the CCS/ONC estimation values are rejected;
or
 - b: liken the estimates of missing people to forms received with very little data and for which EDIS will impute data i.e. use EDIS based on age/sex and geography to impute other characteristics.
38. **Given that the EDIS system retains its current functionality, GROS assumes, subject to further discussion of the details with ONS, that neither of these two options require any significant extra development work since both systems will be running for either the ONC or for the main Census processing. There is a further hybrid option, to choose between CCS imputation and geographically appropriate data from the EDIS system according to some rule as to which should have preference. This would require some development. We would welcome views on the options available to GROS.**

GROS

24 January 2001

Annex A: Scottish Design Groups

Design Group 'name'	Health Board Areas	Population	Council and part Council Areas	Population
Arid	Argyll & Clyde	426,900	Argyll & Bute	89,980
426,900			Inverclyde	85,400
			Renfrewshire	177,830
			Dumbarton District – Helensburgh part area (27,320) (WD1)*	48,760
			Barrhead part area (ER1)*	24,930
Ayar	Ayrshire & Arran	375,400	East Ayrshire	121,300
629,000			North Ayrshire	139,660
			South Ayrshire	114,440
	Borders	106,300	Scottish Borders	106,300
	Dumfries & Galloway	147,300	Dumfries and Galloway	147,300
Fife	Fife	348,900	Fife	348,900
624,700	Forth Valley	275,800	Clackmannanshire	48,560
			Falkirk	144,110
			Stirling	83,130
Grab	Grampian	525,200	Aberdeen City	213,070
525,200			Aberdeenshire	226,260
			Moray	85,870
Greed	Greater Glasgow	911,200	East Dunbartonshire	109,570
911,200			Glasgow city	619,680
			Eastwood (ER2)*	63,050
			Chryston (NL1)*	16,220
			Rutherglen (SL1)*	56,560
			Clydebank (WD2)*	46,120
Hitos	Highland	208,300	Highland	208,300
668,500	Tayside	389,800	Angus	110,070
			Dundee City	146,690
			Perth & Kinross	133,040
	Orkney	19,550	Orkney	19,550
	Shetland	22,910	Shetland	22,910
	Western Isles	27,940	Western Isles	27,940
Leled	Lothian	773,700	East Lothian	89,570
773,700			City of Edinburgh	450,180
			Midlothian	80,860
			West Lothian	153,090
Luck	Lanarkshire	560,800	Cumbernauld & Kilsyth, Monklands and Motherwell Districts (NL2)*	310,500
560,800			Clydesdale, East Kilbride and Hamilton Districts (SL2)*	250,300
Totals				
5,120,000		5,120,000		5,120,000

* NL = North Lanarkshire; SL = South Lanarkshire; ER = East Renfrewshire; WD = West Dunbartonshire. The numbers '1' or '2' refer to parts 1 and parts 2 of those Council Areas.